Appendix A: model details

0.1 Model structure and priors

We used Stan (Carpenter et al., 2017) as the probabilistic language behind the estimation of our Bayesian models in this study, with `brms` as its R interface (Bürkner, 2017). This language implements the Markov Chain Monte Carlo (MCMC) algorithm using the Hamiltonian Monte Carlo method (HMC) to explore the posterior distribution of the model. Broadly, this algorithm is used to iteratively sample the joint sampling space of the parameters to be estimated in the model, and compute, for each value sampled, its likelihood under some probability distribution previously defined. We run four MCMC chains, each 1,000 iterations long each.

0.2 Considerations on statistical power and sample size

There is little consensus about what approach is adequate for calculating the statistical power of a complex Bayesian model like the one in the present study, for several reasons. A first pitfall, shared with frequentist analysis, is that a closed solution for statistical power calculation is not possible or cannot be computed within reasonable time con- straints. This rules out the use of many available pieces of software that are commonly offered for power analysis, as they commonly only consider the case of simpler models like t-tests, ANOVA, Pearson correlation, or regression (with only fixed effects), or trivial derivations of thereof. The more complicated case of multilevel models is usually not covered, not to mention those with a Bayesian approach.

An alternative way of estimating the statistical power of statistical test is simulation. This consists on simulating multiple datasets in which the hypothesised effect size is present, and fitting multiple instances of the model. The statistical power is derived from the proportion of contrasts that result in the rejection of the null hypothesis across datasets. Although this approach permits the estimation of statistical power in the case of more complex models, it involves costly computations. In the case of Bayesian models, and particularly the one in the present study, such cost can be infeasible. Sampling the posterior of our model took approximately seven days. Running this model, or an equivalent one, across 100 datasets (100 may even be considered too few by many) would take more than a year.

Following J. Kruschke (2014), we considered the precision of our estimates as a proxy to statistical power. In particular, we compared the width of the 95% HDI of the critical regression coefficient (*Exposure × Cognateness*) against some nominal interval width. We decided to use the half the width of the ROPE in the logit scale [-0.025, +0.025], that is, 0.05 as the reference interval width. The width of the fixed regression coefficient of *Exposure × Cognateness* ($\beta =$ -0.014, 95% HDI = [-0.017, -0.011]) was 0.006, around 8.933 times narrower than the reference interval. This indicates that the precision of the posterior 95% HDI of the critical parameter in the model is larger than required.

0.3 Model diagnostics

One way to diagnose the behaviour of HMC is to inspect whether the different MCMC chains (if more than one) have converged to a similar region of the posterior. The Gelman-Rubin diagnostic ($\hat{R}$ or R-hat Gelman & Rubin, 1992) provides a measure of chain convergence by comparing the variance within each chain *versus* the variance between each chain. Both are expected to be identical when chains have perfectly converged, so that $\hat{R} = 1$. Values lower than 1.01 are recommended, while values higher than 1.05 indicate that chains might have trouble converging and therefore the estimated parameters must be taken with caution. Figure 1 (A) shows the distribution of $\hat{R}$ values for the coefficients of the fixed effect of our models, which we used for statistical inference. Most values are lower than 1.01, and never higher than 1.05, which provides evidence of successful MCMC convergence.

Another diagnostic of good MCMC converge is the ratio of effective sample size to total sample size (N_{eff}/N), which indicates the proportion of samples in the chain that resulted from a non-divergent transition. Values closer to 1 are ideal, as they indicate that all posterior samples from the MCMC were used to estimate the posterior distribution of the parameter. Values larger than 0.1 are recommended. Figure 1 (B) shows the distribution of the effective sample sizes of the coefficients of the fixed effects in our models. Most values are larger than 0.1, although model 0 ($\mathcal{M}_0$) accumulates most effective sample sizes close to 0.1.

Another way of assessing the behaviour of the HMC algorithm is to visualise the joint posterior distribution for pairs of parameters using bi-variate scatter plots. In Figure 2 we show the pair-wise distribution of posterior samples. Broadly, posterior samples of two parameters should not be correlated. This is the case for all pairs of parameters but for the two intercepts. This is expected behaviour, given that these two parameters correspond to the thresholds between categories in the ordinal regression model, and the distance between both thresholds is fixed in the particular parametrisation of the model.

Finally, we also assessed the predictive performance of the model by doing posterior-predictive checks (PPCs). this involves simulating new datasets from the model and the posterior distribution of its parameters, and checking that,
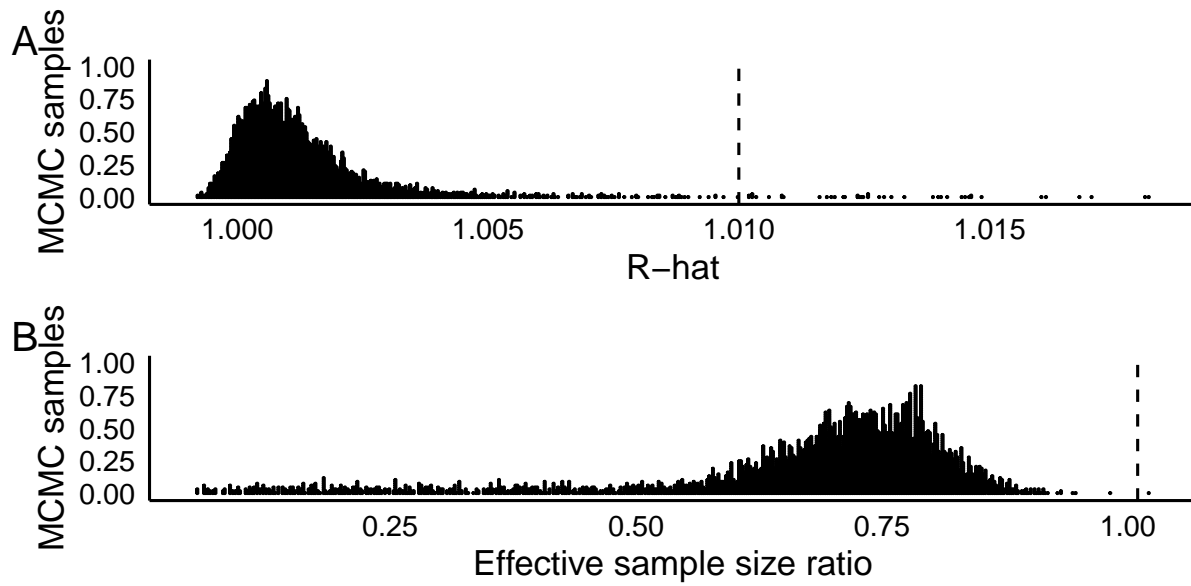
Figure 1: MCMC convergence diagnostic of all parameters in the model. Each dot represents the score of one parameter. (A) Distribution of the Gelman-Rubin (R-hat) scores. (B) Distribution of the ratio of effective sample size.

overall, the distribution of the response variable across the simulated datasets is equivalent to the one in the observed dataset. Figure 3 shows the PPCs of the model.
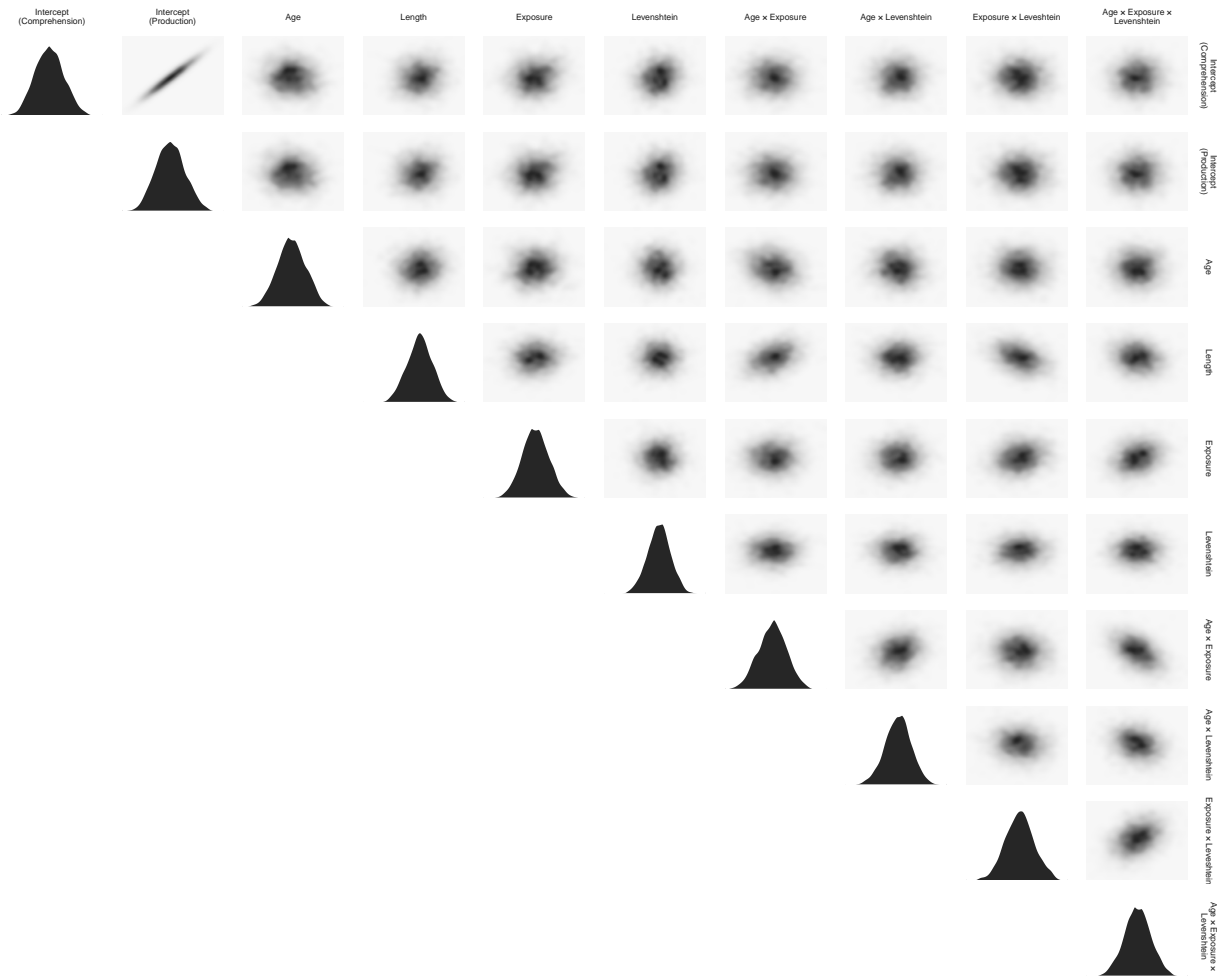
Figure 2: Marginal distribution and bi-variate scatterplot of posterior samples for the fixed regression coefficients in Model 3.
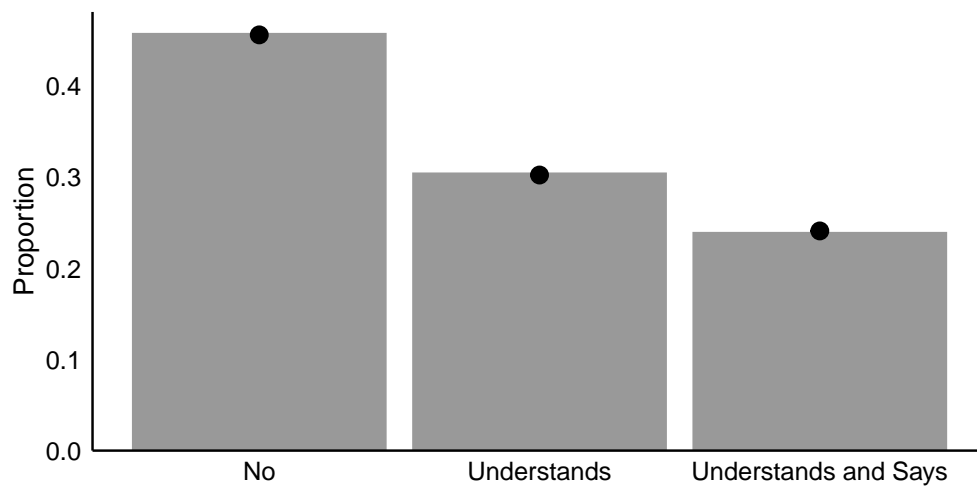

Figure 3: Model posterior predictive checks (PPC). Bars indicate the observed proportion of responses to each category (No, Understands, and Understands and Says). Blue dots and error bars represent the mean proportion of responses simulated from the posterior for each category, and its 95% interval.

Table 1: Posterior distribution of regression coefficients of the model including the *Exposure* composite predictor, and of the model including lexical frequency (*Frequency*) and degree of exposure (*DoE*) separately.

	β	95% HDI	p(ROPE)
Model: Exposure			
Length (+1 SD, 1.56 phonemes)	0.485	[0.478, 0.491]	.000
Age (+1 SD, 4.87 months)	0.600	[0.588, 0.611]	.000
Exposure (+1 SD, 1.81)	0.558	[0.549, 0.566]	.000
Cognateness (+1 SD, 0.26)	0.514	[0.504, 0.526]	.102
Exposure $\times$ Cognateness	0.486	[0.483, 0.489]	.000
Age $\times$ Exposure	0.518	[0.51, 0.526]	.000
Age $\times$ Cognateness	0.504	[0.5, 0.507]	.928
Age $\times$ Exposure $\times$ Cognateness	0.495	[0.493, 0.497]	.908
Model: Frequency & DoE			
Age (+1 SD, 4.87, months)	0.600	[0.59, 0.612]	.000
Phonemes (+1 SD, 1.56 phonemes)	0.486	[0.48, 0.492]	.000
Frequency (+1 SD, 0.19)	0.527	[0.516, 0.539]	.000
DoE (+1 SD, 0.3)	0.557	[0.549, 0.566]	.000
Cognateness (+1 SD, 0.26)	0.516	[0.505, 0.527]	.064
Age $\times$ DoE	0.518	[0.51, 0.526]	.000
DoE $\times$ Cognateness	0.486	[0.483, 0.488]	.000
Age $\times$ Cognateness	0.504	[0.501, 0.507]	.852
Age $\times$ DoE $\times$ Cognateness	0.495	[0.493, 0.497]	.900

β: median of the posterior distribution in the probability scale. 95% HDI: 95% highest density interval of the distribution. p(ROPE): overlap between the 95% HDI and the ROPE, indicating the posterior probability that the true value of the coefficient is equivalent to zero.

Appendix B: frequency and language exposure as separate predictors

As a robustness check, we fit a model similar to the one described in the main manuscript, but including lexical frequency and language degree of exposure as separate predictors, instead of the composite measure *Exposure*. Language degree of exposure (*DoE*) was included in interaction with *Age* and *Cognateness*, while lexical frequency (*Frequency*) was included as a main effect. Table 1 shows a comparison between the posterior distribution of the regression coefficients of both models. Overall, results are equivalent.

Table 2: Sample of items included in the BVQ questionnaire and their syllabified SAMPA transcriptions in Catalan and Spanish.

Translation	Item	X-SAMPA	Syllables	Item	X-SAMPA	Syllables
subway	metro	"mE.t4u	2	metro	"me.t4o	2
glasses	ulleres	u"Le.4@s	3	gafas	"ga.fas	2
house	casa	"ka.z@	2	casa	"ka.sa	2
zipper	cremallera	k4@.m@"Le.4@	4	cremallera	k4e.ma"Le.4a	4
police	policia	pu.5i"si.@	4	policía	po.li"Ti.a	4
ground	sorra	"so.r@	2	arena	a"4e.na	3
white	blanc	b5aN	1	blanco	"blan.ko	2
owl	mussol	mu"sO5	2	búho	"bu.o	2
hammer	martell	m@r"teL	2	martillo	ma4"ti.Lo	3
party	festa	"fes.t@	2	fiesta	"fjes.ta	2
chick	pollet	pu"LEt	2	pollito	po"Li.to	3
ball	pilota	pi"5O.t@	3	pelota	pe"lo.ta	3
doll	nina	"ni.n@	2	muñeca	mu"Je.ka	3
sky	cel	sE5	1	cielo	"Tje.lo	2
pig	porc	pOrk	1	cerdo	"Te4.Do	2

Appendix C: frequency and language exposure as separate predictors

We define syllable frequency as the rate of appearance of individual syllables in the word-forms included in the [QUESTIONNAIRE NAME REDACTED FOR BLIND REVIEW] [REFERENCE REDACTED FOR BLIND RE-VIEW]. Each item corresponds to a Catalan or Spanish word, and has an associated phonological transcription in X-SAMPA format (Wells, 1995). These transcriptions are syllabified. Some examples:

Most Catalan and Spanish words had two syllables, with Spanish words having three and four syllables more often than Catalan words. Less than 1% of the words included in the analyses presented in the main body of the manuscripts had five syllables. No words had more than five syllables (see Figure 4). We extracted lexical frequencies from the English corpora in the CHILDES database (MacWhinney, 2000; Sanchez et al., 2019). Using the Catalan and Spanish corpora was not possible due to the low number of children and tokens included in the corpora.
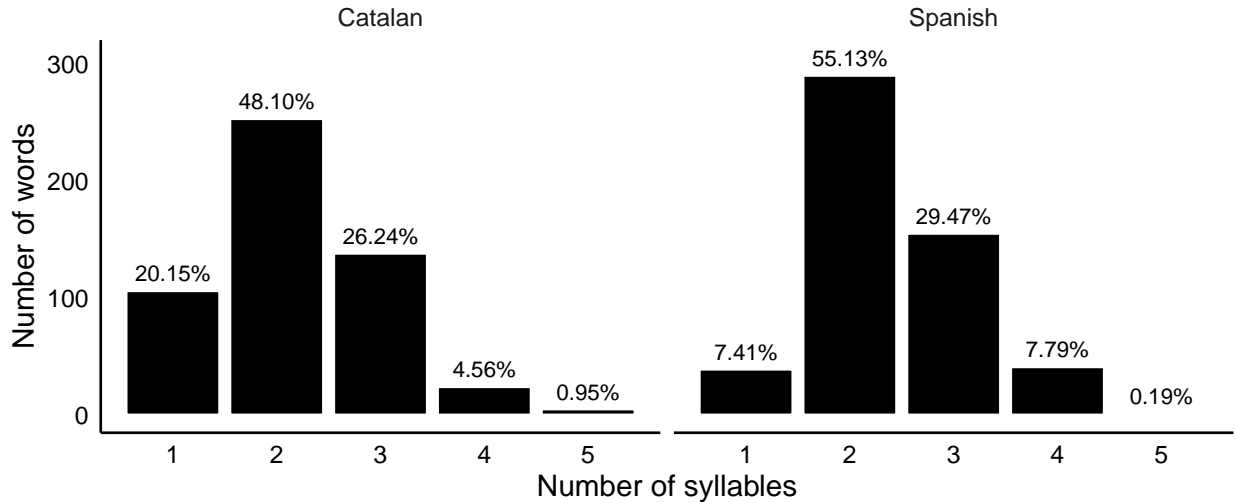

Figure 4: Distribution of the number of syllables in Catalan and Spanish

We now present how syllable frequencies were calculated. Every exposure to a word-form also counts as a exposure to each of the syllables that make up such word. Every time a child hears the word *casa* [house], they are exposed to the syllables *ca* and *sa*. Syllables that appear embedded in words with higher lexical frequency will also be more frequent. To compute the relative frequency of each syllable in Catalan and Spanish (i.e., how many times the syllables appears

75 in every million words in Catalan or Spanish speech), we summed the relative lexical frequency in CHILDES of every
76 word that contains such syllable in the corresponding language. Figure 5 shows the distribution of frequencies across
77 syllables in Catalan and Spanish. In the log10 scale, syllable frequencies in Catalan and Spanish followed a slightly
78 asymmetric distribution, with most syllables scoring around 1,000 counts per million, and a longer tail to the right of
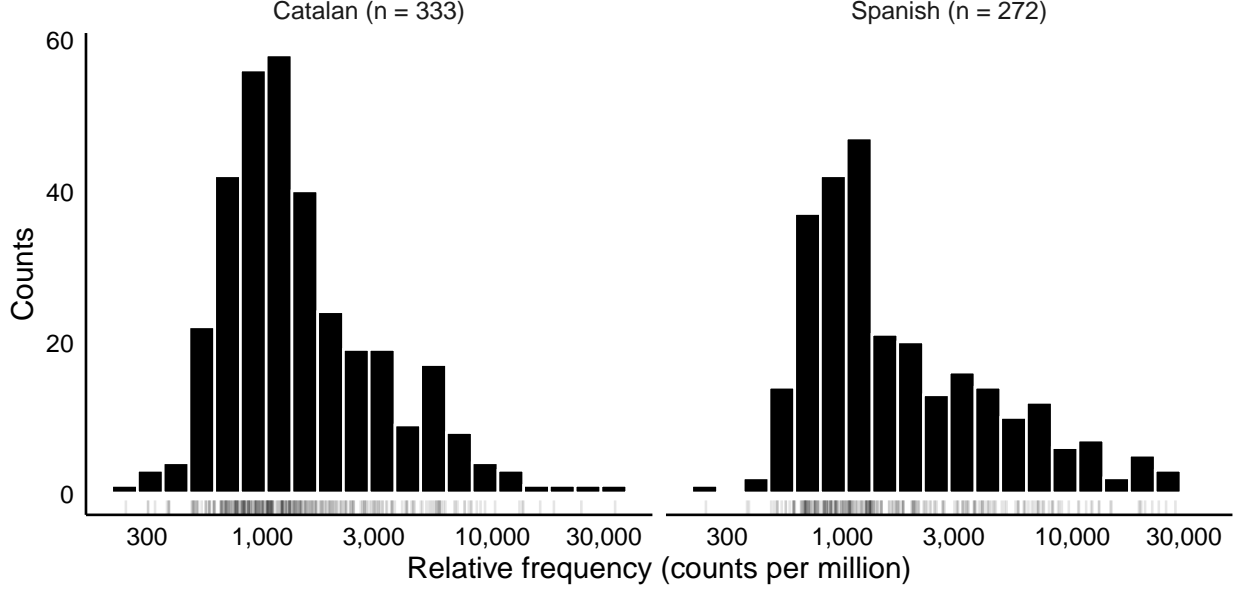79 the distribution.

Figure 5: Distribution of apositional syllable frequencies in Spanish and Catalan

80 To estimate the association between word-level syllabic frequency and cognateness, while controlling for the number
81 of syllables in the word, (an increment in the number of syllables necessarily increases the summed syllabic frequency
82 of the word), we fit a multilevel, Bayesian linear regression model with syllabic frequency (the sum of the syllabic
83 frequency of the syllables in a word) as response variable, and the main effect of the number of syllables ($Syllables$)
84 and $Cognateness$ (Levenshtein similarity between a word and its translation equivalent, Levenshtein, 1966) as pre-
85 dictors. We added translation equivalent-level random effects for the intercept and the main effect of $Syllables$ (some
86 translation pairs had a different number of syllables in each language). We used a Gaussian distribution to model
87 syllabic frequency scores after standardising this variable and the predictors. We used a weakly informative prior for
88 all parameters involved in the model (see Equation 1 for a formal equation of this model and its prior). We conducted
89 statistical inference by evaluating the proportion of the 95% highest density interval (HDI) of the posterior posterior
90 distribution of each coefficient that fell into the region of practical equivalence (ROPE, see the main manuscript for a
91 more detailed explanation, J. K. Kruschke & Liddell, 2018).

$$
\begin{aligned}
\text{Syllable frequency} &\sim \mathcal{N}(\mu, \sigma) \\
\mu &= (\beta_0 + u_{0_i}) + (\beta_1 + u_{1_i})\text{Syllables} + \beta_2\text{Cognateness} \\
\beta_{0-3} &\sim \mathcal{N}(0, 10) \\
u_{0-1_i} &\sim \mathcal{N}(0, \sigma_{u_i}) \\
\sigma_y &\sim \text{Exponential}(2) \\
\begin{pmatrix} u_{0_i} \\ u_{1_i} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u\right) \\
\Sigma_u &= \begin{pmatrix} \sigma_{u_0} & \rho_{u_0}\sigma_{u_0}\sigma_{u_1} \\ \rho_{u_1}\sigma_{u_1}\sigma_{u_0} & \sigma_{u_1} \end{pmatrix} \\
\sigma_{u_{0-1}} &\sim \mathcal{N}_+(1, 0.1) \\
\rho_u &\sim LKJcorr(2)
\end{aligned}
\tag{1}
$$

92 We fit this model running 4 sampling chains with 1,000 iterations each. Table 3 shows a summary of the posterior
93 distribution of the fixed effects in the model. As expected, words with more syllables scored higher in syllabic fre-

Table 3: Posterior distribution of regression coefficients.

	β	95% HDI	$p(\text{ROPE})$
Intercept	16.090	[16.022, 16.162]	NA
Syllables (+1 SD, 0.802)	5.644	[5.575, 5.713]	.000
Cognateness (+1 SD, 0.24)	0.009	[-0.056, 0.081]	1.000

β: median of the posterior distribution in the
probability scale. 95\% HDI: 95\%
highest density interval of the distribution.
$p(\text{ROPE})$: overlap between the
95\% HDI and the ROPE, indicating the posterior
probability that the true value of the coefficient is equivalent to zero.

quency: all posterior draws for the regression coefficient of the main effect of this predictor fell outside the ROPE
defined between -0.5 and +0.5 (β = 5.64, 95% HDI = [5.58, 5.71]). Keeping the number of syllables constant, the
effect of cognateness was negligible: all of the posterior distributions of this predictor fell within the ROPE, providing
evidence that the true value of the increment in syllabic frequency for every increase in cognateness is equivalent to
zero (β = 0.01, 95% HDI = [-0.06, 0.08]).

Figure 6 shows the median posterior-predicted syllabic frequencies for words with one to four syllables, for the whole
range of cognateness values. Overall, cognate words' syllabic frequency is equivalent to that of non-cognates. This
suggests that the cognate facilitation effect in word acquisition reported in the present study is not the result from an
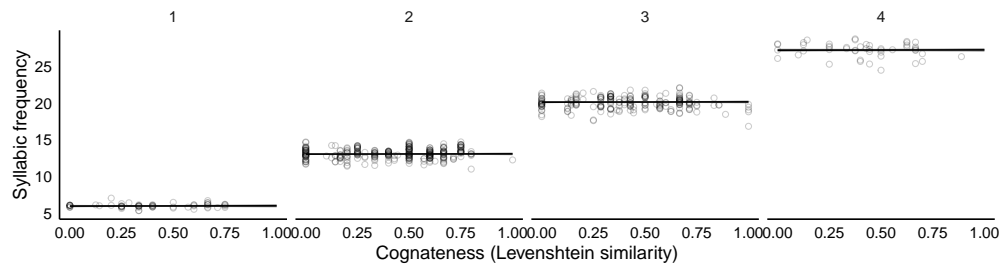association between cognateness and higher syllabic frequencies.

Figure 6: Posterior-predictions of the syllabic frequency model. Thicker lines indicate the median of the posterior
predictions, and thinner lines indicate individual posterior predictions.

References

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan : A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://www.jstor.org/stable/2246093

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, JAGS, and stan*.

Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and planning from a bayesian perspective. *Psychonomic Bulletin &Review*, *25*, 178–206. https://doi.org/10.3758/s13423-016-1221-4

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady*, *10*, 707–710.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). Childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, *51*(4), 1928–1941.

Wells, J. C. (1995). *Computer-coding the IPA: A proposed extension of SAMPA*. *4*(28), 1995.